A hybrid particle-ensemble Kalman filter for problems with medium nonlinearity

Ian Grooms[¶*], Gregor Robinson[¶]

Department of Applied Mathematics, University of Colorado, Boulder, CO, USA

¶These authors contributed equally to this work.
* ian.grooms@colorado.edu

Abstract

A hybrid particle ensemble Kalman filter is developed for problems with medium non-Gaussianity, i.e. problems where the prior is very non-Gaussian but the posterior is approximately Gaussian. Such situations arise, e.g., when nonlinear dynamics produce a non-Gaussian forecast but a tight Gaussian likelihood leads to a nearly-Gaussian posterior. The hybrid filter starts by factoring the likelihood. First the particle filter assimilates the observations with one factor of the likelihood to produce an intermediate prior that is close to Gaussian, and then the ensemble Kalman filter completes the assimilation with the remaining factor. How the likelihood gets split between the two stages is determined in such a way to ensure that the particle filter avoids collapse, and particle degeneracy is broken by a mean-preserving random orthogonal transformation. The hybrid is tested in a simple two-dimensional (2D) problem and a multiscale system of ODEs motivated by the Lorenz-'96 model. In the 2D problem it outperforms both a pure particle filter and a pure ensemble Kalman filter, and in the multiscale Lorenz-'96 model it is shown to outperform a pure ensemble Kalman filter, provided that the ensemble size is large enough.

1 Introduction

Data assimilation of high-dimensional dynamical systems routinely falls to various kinds of ensemble Kalman filters (EnKF) [1]. Ensemble Kalman filters make two fundamental approximations: the first is that the likelihood and prior are both Gaussian, and the second is that the mean and covariance of the Gaussian prior are approximated from an ensemble. The EnKF is known to converge to the correct posterior in the limit of large ensemble size when the distributions are Gaussian [2], but it clearly will not converge to the correct posterior in the presence of non-Gaussianity.

In contrast, Sequential Importance Sampling with Resampling (SIR a.k.a. Particle Filtering) is known to weakly converge to the correct posterior in the large-ensemble limit — with remarkably mild constraints on the dynamics, prior, and observing system [3–5]. This flexibility makes SIR superficially attractive for applications like weather forecasting where nonlinear fluid dynamics lead to non-Gaussian distributions. Unfortunately, however, SIR suffers a severe curse of dimensionality that has prevented its practical application to high dimensional data assimilation problems [6–8]. A variety of methods have been proposed to improve the performance of particle filters in high-dimensional problems, including implicit particle filters [9–11], the equivalent-weights particle filter [12–16], likelihood approximations [17], local particle

filters [18–20] and particle filters based on kernel mappings [21] and synchronization methods [22]. Particle filters have also been hybridized with EnKFs [23–26] and with variational methods [27]. Methods have also been proposed to mitigate the assumption of Gaussianity within the Kalman filter, including nonlinear transformations on the univariate marginal distributions (termed 'Gaussian anamorphosis' in the literature [28–32]) and methods based on rank statistics [33–36].

Although nonlinear dynamics, nonlinear observation operators, and non-Gaussian error distributions lead to non-Gaussian priors and likelihoods in many applications, the degree of non-Gaussianity is not always so great that it severely degrades EnKF performance. This has led several authors to classify problems according to the degree of nonlinearity, i.e. the degree of non-Gaussianity [34, 37, 38]. Following [38] we distinguish three categories:

- Mild nonlinearity: The prior and posterior are both approximately Gaussian.
- Medium nonlinearity: The prior is very non-Gaussian but the posterior is approximately Gaussian.
- Strong nonlinearity: The prior and posterior are both very non-Gaussian.

Particle filters and non-Gaussian extensions of the EnKF are not needed in situations with mild nonlinearity, while problems with strong nonlinearity can greatly benefit from such methods. Problems with medium nonlinearity can arise when nonlinear dynamics produce a non-Gaussian prior, but a highly accurate Gaussian likelihood generates a nearly Gaussian posterior. The concept of medium nonlinearity is related to the Laplace approximation [39]. Morzfeld and Hodyss [38] argue that variational methods are more appropriate for medium nonlinearity than EnKF methods because the former make a Gaussian approximation of the posterior, while the latter make a Gaussian approximation of the prior.

The goal of the present work is to develop a hybrid of the SIR particle filter with the EnKF that is appropriate for problems with medium nonlinearity. The hybrid is based on the likelihood splitting of Frei & Künsch [23]. At each assimilation cycle, part of the observational information is incorporated by means of an SIR step, and then the remaining observational information is incorporated with a serial square root version of the EnKF. Particle degeneracy that results from the resampling step of the SIR is broken by a mean preserving random orthogonal transformation of the ensemble, as seen in certain EnKFs [40–42] and moment-matching particle filters [43, 44]. The goal of the hybrid is to present the EnKF with an intermediate prior that is closer to Gaussian than the true prior. The curse of dimensionality in the particle filter is mitigated by assimilating only part of the observational information, i.e. only moving partway from the prior to the posterior, thereby enabling accurate results with practical ensemble sizes. The hybrid presented here is broadly similar to other hybrids (e.g. [23, 24]), and differs mainly in the explicit focus on problems with medium nonlinearity and in details of the implementation. Differences are discussed further in section 2.3.

The hybrid particle ensemble Kalman filter is presented in section 2. The new hybrid is compared to the hybrids from [23, 24] and to a particle filter and an ensemble Kalman filter in the context of a simple two-dimensional problem in section 3. A multiscale Lorenz-'96 model from [45] is described in section 4.1, followed by a description of the data assimilation system configuration in section 4.2. The EnKF component of the hybrid uses multiplicative inflation and localization, and the method used to optimize the values of these parameters is described in section 4.3. Results of the experiments are described in section 5, followed by a conclusion in section 6.

2 The hybrid algorithm

2.1 SIR

Standard sequential importance resampling (SIR) particle filters work as follows [3, 46]. Each ensemble member $\boldsymbol{x}_0^{(i)}$ (or 'particle') starts with equal weight $w_0^{(i)} = 1/N$, where N is the ensemble size and $i = 1, \ldots, N$. Subscripts refer to time, while superscripts in parenthesis refer to ensemble members. Each ensemble member is forecast until the next assimilation cycle. At assimilation cycle j the weights are updated using the likelihood $L(\boldsymbol{x})$

$$w_j^{(i)} = \tilde{w}_{j-1}^{(i)} \frac{L\left(\boldsymbol{x}_j^{(i)}\right)}{Z_j} \tag{1}$$

where Z_j is a normalization constant to ensure that the weights sum to one, and $\tilde{w}_j$ denotes the effect of resampling: without resampling at step j we have $\tilde{w}_j^{(i)} = w_j^{(i)}$ whereas with resampling we have $\tilde{w}_j^{(i)} = 1/N$. A resampling is then applied whereby particles with high weights are replicated and particles with low weights are eliminated. There are a variety of resampling algorithms; here we use the so-called 'systematic' resampling scheme of [47].

It is well known that the weights of a particle filter can collapse, especially in high dimensions, i.e. a small number of particles receive a weight near one while all others receive a weight near zero [7, 8]. After resampling, only the high-weight particles are left. If an optimal-transport based alternative to resampling is used [24, 48, 49], then all particles are transported to a very small vicinity of the high-weight particles. In both cases the posterior distribution is poorly estimated. The number of particles with a substantial portion of the weight can be approximated by the effective sample size

$$\mathrm{ESS} = \frac{1}{\sum_{i=1}^{N} \left(w_j^{(i)}\right)^2}. \tag{2}$$

The ESS takes values between 1 and N, and small ESS indicates that the weights have collapsed.

2.2 Ensemble Square Root Filter (ESRF)

There are many ensemble Kalman filters, any of which could be hybridized with the smoothed particle filter. We focus here on an ensemble square root filter (ESRF) developed in [50] for sequential assimilation of observations possessing uncorrelated errors. At a single assimilation cycle the ensemble is denoted $\{\boldsymbol{x}^{(i)}\}_{i=1}^{N}$. The ensemble mean is denoted $\bar{\boldsymbol{x}}$, and the scaled ensemble perturbation matrix is denoted

$$\mathbf{A} = \frac{1}{\sqrt{N-1}} \left[\boldsymbol{x}^{(1)} - \bar{\boldsymbol{x}}, \ldots, \boldsymbol{x}^{(N)} - \bar{\boldsymbol{x}}\right]. \tag{3}$$

The ensemble covariance matrix is thus $\mathbf{A}\mathbf{A}^T$. Covariance inflation is applied by replacing $\mathbf{A}$ with $\sqrt{1+r}\mathbf{A}$, where $r > 0$ is a tunable inflation factor.

Observations are linear, and a single scalar observation y takes the form

$$y = \boldsymbol{h}^{\boldsymbol{T}}\boldsymbol{x} + \epsilon. \tag{4}$$

Here the observation error ϵ is a sample from a zero-mean normal distribution with variance γ^2 and the row vector $\mathbf{H} = \boldsymbol{h}^T$ extracts the observations from the state vector

$\boldsymbol{x}$. It is convenient to define the row vector $\boldsymbol{v}^T = \boldsymbol{h}^T\mathbf{A}$. With this notation, the ESRF from [50] corresponds to the following update of the ensemble mean

$$\bar{\boldsymbol{x}}^a = \bar{\boldsymbol{x}} + \frac{(y - \boldsymbol{h}^T\bar{\boldsymbol{x}})}{\sigma^2 + \gamma^2}\mathbf{A}\boldsymbol{v} \tag{5}$$

and the following update of the scaled ensemble perturbation matrix

$$\mathbf{A}^a = \mathbf{A} - b\mathbf{A}\boldsymbol{v}\boldsymbol{v}^T, \tag{6}$$

$$b = \frac{1}{\sigma^2 + \gamma^2 + \gamma\sqrt{\sigma^2 + \gamma^2}} \tag{7}$$

where $\sigma^2 = \boldsymbol{v}^T\boldsymbol{v}$ and $\mathbf{A}^a$ is the scaled analysis ensemble perturbation matrix.

Localization is applied by multiplying the increments elementwise by a localization vector $\boldsymbol{\rho}$. The elements of $\boldsymbol{\rho}$ are $e^{-(d/L)^2/2}$, where d is the distance from $\boldsymbol{x}_i$ to y and L is a tunable localization radius. This amounts to updating Eq 5 and Eq 6 to

$$\bar{\boldsymbol{x}}^a = \bar{\boldsymbol{x}} + \frac{(y - \boldsymbol{h}^T\bar{\boldsymbol{x}})}{\sigma^2 + \gamma^2}\boldsymbol{\rho} \circ (\mathbf{A}\boldsymbol{v}) \tag{8}$$

and

$$\mathbf{A}^a = \mathbf{A} - b\left(\boldsymbol{\rho} \circ (\mathbf{A}\boldsymbol{v})\right)\boldsymbol{v}^T \tag{9}$$

where $\circ$ denotes an elementwise product.

Evensen was the first to suggest resampling the posterior within the context of an ensemble square root filter by multiplying $\mathbf{A}^a$ from the right by a random orthogonal matrix [40]. Since the posterior ensemble covariance matrix is $\mathbf{A}^a\mathbf{A}^{aT}$, this kind of resampling does not change the ensemble covariance matrix. Sakov & Oke [42] pointed out that the random orthogonal matrix should have $\mathbf{1}$ (the vector whose elements are all 1) as an eigenvector in order for the resampling to preserve the ensemble mean. We construct a new scaled ensemble perturbation matrix $\mathbf{A}^a$ by multiplying $\mathbf{A}^a$ from the right by a random orthogonal matrix $\mathbf{Q}$ that has $\mathbf{1}$ as an eigenvector. The matrix $\mathbf{Q}$ is constructed as follows [42]

$$\mathbf{Q} = \mathbf{U}\left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{P} \end{array}\right]\mathbf{U}^T. \tag{10}$$

The matrix $\mathbf{U}$ is an orthogonal matrix whose first column is proportional to $\mathbf{1}$, while the matrix $\mathbf{P}$ is a random orthogonal matrix of size $N-1 \times N-1$. The matrix $\mathbf{U}$ is time independent. With a large ensemble size it can become costly to sample a new $\mathbf{P}$ at each assimilation cycle. In principle the matrix $\mathbf{Q}$ could be constructed once and used repeatedly, but in our numerical experiments $\mathbf{P}$ is resampled at each assimilation cycle.

Using this method, a single assimilation cycle proceeds as follows

- Form the ensemble mean $\bar{\boldsymbol{x}}$ and scaled ensemble perturbation matrix $\mathbf{A}$.
- Inflate the scaled ensemble perturbation matrix: $\mathbf{A} \leftarrow (1 + r)\mathbf{A}$
- For each observation, find $\bar{\boldsymbol{x}}^a$ and $\mathbf{A}^a$ using Eq 8 and Eq 9.
- Resample the posterior ensemble by replacing $\mathbf{A}^a$ with $\mathbf{A}^a\mathbf{Q}$.
- Reconstitute the ensemble according to $\boldsymbol{x}^{(i)} = \bar{\boldsymbol{x}}^a + \sqrt{N-1}\mathbf{A}_i$ where $\mathbf{A}_i$ is the i^{th} column of $\mathbf{A}$.

2.3 SIR-ESRF hybrid

The SIR/ensemble square root filter (SIR-ESRF) hybrid developed here is based on the bridging method of Frei and Künsch [23]. The likelihood $L(\boldsymbol{x})$ is split into a product $(L(\boldsymbol{x}))^{\alpha} \cdot (L(\boldsymbol{x}))^{1-\alpha}$ where $\alpha \in [0,1]$ is the "splitting factor". The hybrid proceeds by having the SIR particle filter assimilate using the likelihood $(L(\boldsymbol{x}))^{\alpha}$, followed by an ESRF assimilation using the likelihood $(L(\boldsymbol{x}))^{1-\alpha}$. In principle, the methods can be applied in either order [24], but the method developed here is intended for situations where the prior is non-Gaussian but the posterior is nearly Gaussian ('medium' nonlinearity according to [38]). In such cases the intermediate posterior produced after the first assimilation with the particle filter should be closer to Gaussian than the prior. The ESRF subsequently performs an assimilation on a problem that more closely conforms to its underlying Gaussian approximation.

Following Frei & Künsch [23] we choose the splitting factor α to ensure that the effective sample size is within some tolerance of a tunable theshold. This is achieved with a rootfinding method. A large ESS threshold implies a small α, though the precise value of α depends on the ensemble size. If $\alpha = 0$, then the hybrid reverts to a pure ESRF because all the particle filter weights become equal.

The resampling step of the SIR particle filter leads to a degeneracy where there are multiple copies of some ensemble members. In our numerical experiments we use a deterministic system of ordinary differential equations, so the dynamics do not break the degeneracy. We opt to follow the ESRF assimilation with a mean-preserving random orthogonal transformation that resamples the ensemble within the Gaussian posterior, as described in the foregoing section.

There are two other extant hybrid particle/ensemble Kalman filters: those of [23] and [24]. Our hybrid is essentially the same as the hybrid of [24] with the following differences: We use standard resampling methods for the particle filter part of the hybrid instead of the Ensemble Transform Particle Filter (ETPF) method of [48], and we break degeneracy using a random orthogonal transformation rather than the 'particle rejuvenation' procedure of [24]. Our use of a random orthogonal transformation is motivated by the focus on medium nonlinearity problems. Naive implementations of the ETPF are computationally expensive, and in the experiments with the Hénon map described in section 3 there seems to be little benefit in using the ETPF instead of standard resampling.

The hybrid of [23] is significantly different from the one proposed here and from the hybrid of [24] because the first step of Frei & Künsch's hybrid is really a Gaussian mixture model update and not a particle filter update (cf. [51]), though it does limit to a pure SIR particle-filter update in the limit $\alpha \to 1$. In particular the particle weights in the hybrid of [23] are different from those used here and in [24], and are more expensive to evaluate. Particle degeneracy is avoided in the hybrid of [23] by using a stochastic update for each step: a perturbed-observation Gaussian mixture update (cf. [51]) for the first step and a perturbed-observation EnKF for the second step (cf. [52, 53]). It is worth noting that the hybrids of [23] and [24] are generally intended to overcome non-Gaussianity in the filtering problem. The hybrid developed here is quite similar to that of [24] but has a tighter focus: we expect the hybrid to achieve near-optimal performance on problems with medium nonlinearity, but not on problems with strong nonlinearity.

2.4 Blurring observations

The development of particle filters that avoid or reduce the incidence of collapse is an active area of research. The authors recently proposed an alternative that uses the same forecast as the standard particle filter, but imposes a generalized random field model of

observation errors [17]. When the observation errors are Gaussian, the likelihood takes the form

$$L(\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x}))^T\mathbf{R}^{-1}(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x}))\right\}. \tag{11}$$

Here $\boldsymbol{y}$ is the observation vector, $\boldsymbol{H}$ is the observation (or 'forward') operator, and $\mathbf{R}$ is the observation error covariance matrix. In the particle filter of [17], the observation error covariance matrix $\mathbf{R}$ is replaced by a covariance matrix that has increasing variance at small spatial scales. In practice this is implemented by blurring (i.e. smoothing) the innovations $\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x})$. The authors recently developed a fast algorithm for blurring scattered data in arbitrary dimensions for this purpose [54].

In the numerical experiments presented here, the spatial domain is periodic and Fourier methods are used to apply the blurring. The true observation error covariance matrix is $\mathbf{R}=\gamma^2\mathbf{I}$. In the particle filter with blurred observations this is replaced by $\gamma^2\left(\mathbf{S}^T\mathbf{S}\right)^{-1}$, where the matrix $\mathbf{S}$ corresponds to an operator that attenuates the Fourier coefficients using the following spectrum

$$\frac{1}{(1+(\ell k)^2)^\beta} \tag{12}$$

where β and ℓ are tunable parameters and k is the Fourier wavenumber. More general blurring spectra are trivial to implement in our experiments, but the above blurring corresponds to the spectrum of the fast algorithm for scattered data developed in [54].

Replacing the true likelihood by a likelihood associated with spatial blurring means that the particle filter is approximating a distribution other than the true Bayesian posterior. The effect of this blurring is to make the likelihood uninformative at small scales, so that the posterior reverts to the prior at small scales. At large scales the blurring likelihood is close to the true likelihood, so the approximate posterior is close to the true posterior. Blurring reduces the effective dimension of the problem by confining the dimensionality to that of the large scales. This has the effect of reducing the minimum ensemble size needed to avoid collapse. It can also improve uncertainty quantification of large scales for a fixed ensemble size.

3 Numerical experiment: Hénon map

This section serves to illustrate a specific problem with medium nonlinearity, and to compare the three hybrid particle/ensemble Kalman filters with a particle filter and an ensemble Kalman filter. Rather than performing a cycled data assimilation experiment where the output of one cycle serves as the initial condition for the next, we repeat the same experiment multiple times. This serves to focus attention on a single Bayesian assimilation update, avoiding the complication associated with cycled data assimilation where the degree of non-Gaussianity can vary from one cycle to the next.

The prior imposed is the joint distribution of U and V obtained by applying one iteration of the Hénon map to a standard normal initial condition on U_0 and V_0, i.e.

$$\begin{aligned} U &= 1-1.4U_0^2+V_0, \\ V &= 0.3U_0. \end{aligned}$$

The prior probability density is shown in color in the upper left panel of Fig 1. The true values of U and V are set to -4 and 0.6, respectively, and the observation is drawn from the normal distribution with mean equal to the true value of U and V and diagonal covariance with entries 1 and 0.01. The resulting posterior probability distribution is approximately Gaussian, as shown by the contours in the upper left panel

Fig 1. Upper left: Prior distribution (color) and posterior distribution (contours). The remaining panels illustrate the five methods. In each panel the blue dots show the prior ensemble, the black dots show the posterior ensemble, and the green dot shows the observation. In the right column, the orange dots represent the intermediate ensemble produced by the first step of each hybrid.

of Fig 1 (where the observation is without error, for convenience). Since the prior is clearly non-Gaussian, a pure EnKF solution is expected to give a biased result regardless of ensemble size. In contrast, the hybrid should achieve nearly optimal performance provided that the ensemble size is large enough to avoid sampling errors.

To illustrate these ideas we compare five methods:

(i) ETPF: A pure particle filter from [48]

(ii) ESRF: A pure ESRF described in section 2.2

(iii) GMM-EnKF: The Gaussian mixture model – EnKF hybrid of [23]

(iv) ETPF-ESRF: The hybrid of [24] combining the ETPF and the serial square root ESRF described in section 2.2

(v) SIR-ESRF: The hybrid described in section 2.3 that combines a standard SIR particle filter and an ESRF with a mean-preserving random orthogonal resampling

These five methods are illustrated in Fig 1, using an ensemble size of 100; in every panel the blue dots represent the prior sample and the green dot shows the true value of U and V. The black dots represent the posterior ensemble in each panel, and in the panels illustrating the hybrid methods the orange dots represent the sample from the intermediate posterior distribution.

The center left panel illustrates the particle filter (ETPF). The ETPF posterior sample is tightly clustered around a small number of the prior samples, which reflects the fact that the ESS is very low (ESS = 5 in this example), despite having an ensemble size of 100 for a problem with dimension 2. This illustrates the severe ensemble size requirements of particle filters. The lower left panel shows the ESRF. The ESRF produces a posterior close to the true value in this case, but the posterior ensemble it produces shows clear discrepancy from the true posterior.

The hybrids all choose the split α to produce an ESS of 30. ETPF-ESRF and SIR-ESRF are shown in the center right and lower right panels, respectively; they use the same split α and the same particle weights. The two methods produce very similar results; one notable difference is that ETPF-ESRF produces an intermediate distribution with less particle degeneracy than SIR-ESRF. This difference in the intermediate distribution does not have a significant impact on the final posterior distribution. GMM-EnKF (upper right panel) uses a different formula for the particle weights — because the first step is a Gaussian mixture model rather than a sum of delta distributions — and thus chooses a different split α to achieve the target ESS of 30. As a result, GMM-EnKF produces an intermediate distribution that is more tightly clustered on the observation in comparison to the other hybrids. The posterior ensemble is also slightly less dispersed than the other hybrids, but is qualitatively similar.

To carefully compare the performance of the different methods, we solve the problem 1,000 times for each method over a range of ESS thresholds. The results are compared on the basis of the root mean squared error (RMSE) where the mean is taken over the 1,000 experiments, and the continuous ranked probability score (CRPS; [55, 56]). The median of these 1,000 CRPS values is used as a summary statistic. We also run a

Fig 2. Filter performance over 1,000 trials as a function of ESS threshold. Top row: RMSE. Bottom row: Median CRPS. Left column: Results for the U variable. Right column: Results for the V variable. The ETPF results are shown in the plots at ESS = 0 even though the ESS for ETPF is typically around 4.4. The ESRF results are shown in the plots at ESS = 100. The dashed line in each panel shows the performance of an SIR particle filter with 10^4 particles.

standard SIR particle filter with 10^4 particles, as a reference approximation of the true Bayesian posterior.

Figure 2 shows the performance of the methods as a function of the ESS threshold. The top panels show RMSE and the bottom panels show the median of CRPS for the U (left) and V (right) variables. The mean ESS of the ETPF over 1,000 trials is 4.4, so the smallest ESS threshold was set to 10. The ETPF performance is shown on the plots at ESS= 0, purely for convenience. The pure ESRF performance is shown on the plots at ESS= 100. In each panel the performance of the pure particle filter with an ensemble size of 10^4 is shown for reference.

All three hybrids perform similarly, though the hybrid of [23] performs slightly worse than the other two in terms of RMSE. This may be because the first step of the GMM-EnKF hybrid uses a GMM whose component Gaussians all use a covariance matrix obtained from the full prior ensemble; performance might be improved by using a clustering approach in the GMM following [57]. The hybrids of [24] and section 2.3 are both able to perform *better* than the pure particle filter when the ESS threshold is low, and are able to nearly match the performance of the true Bayesian filter as approximated by the pure particle filter with 10^4 particles. This is because the pure particle filter with 100 particles is still limited by low ESS (as underscored by the typical ESS value of 4.4).

The differences in RMSE between the methods are fairly small – on the order of 25% at most. Differences in CRPS, which measures the quality of the uncertainty quantification (UQ) associated with the ensemble, are much larger. The hybrid filters all achieve nearly optimal UQ, achieving more than 50% improvement in CRPS over both ETPF and ESRF.

The pure particle filter is quite general in the sense that it generates a consistent estimator of the true Bayesian posterior for a wide range of problems. The cost of this generality is the requirement of a very large ensemble size. The hybrids trade this generality for improved performance using smaller ensemble sizes on a specific subset of problems, namely those with medium nonlinearity.

The code and data associated with this section can be found in [58].

4 Numerical experiment: Lorenz-'96

4.1 A two-scale Lorenz-'96 Model

The experiments in this section make use of a model inspired by the Lorenz-'96 model [59, 60] and developed in [45]. The standard two-scale (or 'two-layer') Lorenz-'96 model includes two sets of variables, X_k and $Y_{j,k}$. There are fewer X_k variables, and they evolve more slowly than the $Y_{j,k}$ variables, so the X_k variables are typically viewed as 'large-scale' while the $Y_{j,k}$ variables are viewed as 'small-scale.' The difficulty with this model is that it lacks a clear connection to a spatial field of a physical quantity like temperature or velocity, observations of which contain both large and small scales. A model inspired by the Lorenz-'96 models that possesses a single set of variables x_i with distinct large-scale and small-scale dynamics was developed in [45] and has been used

Fig 3. A simulation of the two-scale Lorenz-'96 model initialized at $t = 0$ with a sample from a standard normal distribution.

recently as a test model for data assimilation in [61]. The model is governed by a system of ordinary differential equations of the form

$$\dot{\boldsymbol{x}} = h\boldsymbol{N}_S(\boldsymbol{x}) + J\mathbf{T}^T\boldsymbol{N}_L(\mathbf{T}\boldsymbol{x}) - \boldsymbol{x} + F\mathbf{1} \tag{13}$$

where $h, F \in \mathbb{R}$, $J \in \mathbb{N}$, $\mathbf{1}$ is a vector of ones, and

$$\left(\boldsymbol{N}_S(\boldsymbol{x})\right)_i = -x_{i+1}(x_{i+2} - x_{i-1}) \tag{14}$$

$$\left(\boldsymbol{N}_L(\boldsymbol{X})\right)_k = -X_{k-1}(X_{k-2} - X_{k+1}). \tag{15}$$

The number of state variables in $\boldsymbol{x}$ is $41J$; here $J = 128$ for a total system dimension of 5248. As in the Lorenz-'96 model, the indices extend periodically. The matrix $\mathbf{T}$ projects onto the 41 largest-scale discrete Fourier modes and then evaluates that projection at 41 equally-spaced points on the grid of state variables. The matrix $J\mathbf{T}^T$ interpolates a vector of length 41 back to the full dimension of $\boldsymbol{x}$.

The large-scale part of the model dynamics is obtained by applying $\mathbf{T}$ to $\boldsymbol{x}$. The result is identical to large-scale dynamics of the standard Lorenz-'96 model, except that the large scales are coupled to small scales via the term $h\mathbf{T}\boldsymbol{N}_S(\boldsymbol{x})$. While the Lorenz-'96 model is often configured with 40 large-scale variables (e.g. [62]), [45] used 41 variables so that the 20^{th} Fourier mode is not split between large and small scales. At small scales, the dynamics are the same as those of original Lorenz-'96 model but with the direction of indexing reversed.

The experiments presented here use $h = 0.38$ and $F = 8$. With these parameters the large-scale dynamics are very similar to the standard Lorenz-'96 model, with fairly weak coupling to the small scales. The exception is when the large-scale Lorenz-'96 component reaches large values (e.g. amplitudes ≥ 10). This occurrence excites a fast small-scale instability, causing the small scales also to reach large amplitudes that feed back locally onto the large-scale dynamics. Fig 3 shows the result of a simulation of this model initialized at $t = 0$ with a sample from a standard normal distribution. After a short transient the dynamics settle onto an attractor, with large-scale Lorenz-'96 modes propagating eastward and small-scale instabilities transiently excited by the large-scale waves.

4.2 Data assimilation system configuration

Reference solutions are generated by drawing initial conditions from an uncorrelated standard normal distribution and propagating the initial conditions by 9.0 time units by numerical intergration of the dynamical model, at which point the state arrives at a statistical steady state (cf. Fig 3). Upon reaching that statistically steady state, a reference state is produced at 1500 time intervals separated by 1.2 time units. In the usual interpretation of the standard Lorenz-'96 model, this time interval corresponds to 6 days, which is quite long compared to other studies. At shorter time intervals the model exhibits only mild nonlinearity, where the forecast distribution is still very nearly Gaussian even though the dynamics are nonlinear. At 6 days the forecast distributions are certifiably non-Gaussian, as shown in Fig 4. This figure was produced by projecting a forecast ensemble of 1200 members onto the three leading singular vectors of the ensemble's empirical covariance matrix. The forecast distribution is dramatically non-Gaussian within this subspace — therefore the EnKF assumption of a Gaussian prior is invalid.

Fig 4. After one ensemble forecast (ensemble size is 1200) the deviations from the forecast mean are projected into the three leading eigenvectors of the empirical covariance matrix, with projection coefficients denoted x, y, and z. The four panels show four different perspectives on the projected ensemble. The color of each dot corresponds to the particle filter weight assigned using a split α chosen to yield an effective sample size of ESS = 600.

Our hybrid is intended for situations with medium non-Gaussianity, where the prior is not Gaussian but the posterior is nearly Gaussian. To achieve an approximately Gaussian posterior in the face of a non-Gaussian prior requires a large number of sufficiently-accurate observations. Observations are taken at every fourth grid point (i.e. 32 observations for each of the 41 large-scale modes), with observation error variance $\gamma^2 = 1/2$. This density and accuracy of observations is sufficient to produce a nearly-Gaussian posterior without rendering the data assimilation procedure superfluous. (If the observations are dense enough and accurate enough then the filter adds essentially no information to the observations; this situation is avoided here, as the filter accuracy remains better than the observational accuracy.)

Ensemble members are initialized by propagating a sample from the uncorrelated multivariate standard normal distribution by 9.0 time units to arrive at an ensemble of substantially disparate states near the dynamic's attractor. Because this initial forecast ensemble is fairly uninformative of the true state, there is a transient in filter performance while the filter approaches its asymptotic optimal performance. The results of the first 100 assimilation cycles are ignored in computations of filter performance statistics, so that the results presented are reflective of the statistical steady state of the filter. The data assimilation system was run for 1500 cycles, i.e. nearly 25 years, for each trial in the experiment.

4.3 Parameter Optimization

The ESRF used here has two primary tunable parameters: the inflation factor r and the localization radius L. The SIR-ESRF hybrid filter has an additional tunable parameter, the ESS threshold that determines the splitting factor α. The version of the hybrid filter with blurred observations (denoted BSIR-ESRF) also has tunable parameters related to the blurring, but these should not be viewed as a primary means of optimizing performance; we expect the hybrid to outperform the pure ESRF using only reasonable blurring parameters chosen a priori. The demarcation between large and small scales occurs at Fourier wavenumber 20 for the Lorenz-'96 model considered here, so the blurring is chosen to have a Fourier spectrum

$$\frac{1}{\left(1 + \left(\frac{k}{20}\right)^2\right)^2}.$$

To help substantiate a comparison between our SIR-ESRF hybrid approach and the pure ESRF filter, we independently tuned the respective filter parameters. This began by generating parameter configurations, described hereafter as "arms," from a Sobol sequence of low-discrepancy quasirandom numbers in a bounding box that we chose as a search space [63]. (The term "arm" comes from the literature on multi-armed bandits and denotes a particular configuration to be tested.) The range of inflation factors considered was from $r = 0$ to $r = 0.08$ for the pure ESRF, and from $r = 0$ to $r = 0.15$ for the hybrid. The range of ESS thresholds for the hybrid was from 66 to 400 for $N = 400$ and 200 to 1200 for the $N = 1200$. The range of localization radius L was from 128 points (equal to the separation between large-scale Lorenz-'96 modes) and 320

points. At larger localization radii the filter performance became highly erratic, with some experiments performing extremely well and others extremely poorly. It seems likely that at large localization radii there are rare occurrences of spurious long-range correlations that significantly degrade the filter performance.

For each arm, we ran at least four separate experiments with different reference solutions and initial ensembles. For each assimilation cycle we computed the resulting root mean square error (RMSE), spread, and continuous ranked probability score (CRPS; [55, 56]) for both the forecast (prior) and analysis (posterior). RMSE and spread are scalar quantities at each timestep, but CRPS was computed for each state variable at each timestep. We then aggregated these quantities by computing a mean over all state variables, timesteps, and assimilation trials — excluding the first 100 timesteps to allow for filter burn-in.

We elected to optimize for mean analysis CRPS because it quantifies the accuracy of the entire distributional estimate, whereas RMSE only describes accuracy of the ensemble mean point estimate. The ensemble spread would also provide an estimate of the distributional accuracy, but CRPS is preferable in its ability to quantify the accuracy of non-Gaussian distributional estimates. The median was excluded as an aggregation function to optimize because we found it to be insufficiently sensitive to situations in which the filter produces large intermittent excursion from the true state.

After exploring broad patterns with a Sobol sequence, we switched to a Bayesian optimization method for choosing new arms to evaluate. Using a Bayesian optimization method substantially accelerated convergence to optimal filter parameters relative to the quasirandom search. In short, this involved fitting a Gaussian process surrogate model to the mean CRPS observations as a function on the parameter search space, and then choosing new arms that maximize a utility function under that surrogate model. We chose a utility function that estimates improvement from previously observed arms expected under the surrogate model. The arms are then evaluated in parallel, by running the filter on a subset of the reference simulations using those arms' filtering parameters. Those results are then incorporated with previous results to fit a new Gaussian process surrogate model used in the next iteration of the Bayesian optimization loop. The technical details of the optimization strategy we used are described in S1 Appendix.

5 Results: Lorenz-'96

The three methods have indistinguishable performance at ensemble sizes smaller than 400, and the performance of all three methods improves with increasing N up to $N = 400$. This suggests that for $N < 400$ sampling errors limit filter performance more than errors due to non-Gaussianity. It is probable that this threshold could be reduced with more sophisticated inflation and localization strategies (e.g. [64, 65]). Table 1 presents the results for the optimal parameter configurations of each method at two ensemble sizes $N = 400$ and $N = 1200$.

5.1 $N = 400$

At an ensemble size of 400 the three methods yield very similar results. In all three cases the filter is clearly doing better than simply trusting the observations, because the RMSE is nearly half the standard deviation of observation error. The SIR-ESRF hybrid is slightly worse than the other two on average, because three of the eight runs produced significantly worse results, with analysis CRPS above 1.4. In contrast, the BSIR-ESRF hybrid produces results quite similar to the ESRF. One notable difference is that the optimal inflation parameter r is larger for the hybrid filters than for the pure

		Analysis		Forecast	
N	Configuration	CRPS	RMSE	CRPS	RMSE
400					
	ESRF	0.115	0.296	0.548	1.13
	SIR-ESRF	0.125	0.328	0.554	1.13
	BSIR-ESRF	0.111	0.287	0.533	1.10
1200					
	ESRF	0.112	0.280	0.539	1.11
	SIR-ESRF	0.115	0.281	0.530	1.08
	BSIR-ESRF	0.106	0.266	0.500	1.03

Table 1. Results for optimal parameter configurations of each method: pure ESRF, hybrid SIR-ESRF, and hybrid with blurred observations BSIR-ESRF. Results are averaged over the last 1400 assimilation cycles and over 8 different sets of initial conditions.

ESRF, presumably to counteract the under-dispersion that results from the resampling step in the particle filter. (The optimal r for SIR-ESRF is 0.06 vs 0.026 for ESRF.) The optimal effective sample size for the SIR-ESRF hybrid was 297, which is fairly large compared to the ensemble size of 400.

Figure 5 shows the GP surrogate model's prediction for analysis CRPS as a function of localization radius and inflation ratio $(1 + r)$ for the pure ESRF filter at $N = 400$. Gray squares indicate parameter configurations where experiments were run. The left panel plots the mean of the GP, while the right panel plots the standard deviation. The optimal parameters are in a fairly broad well, with near-optimal localization radii ranging from 100 to 300 and corresponding inflation factors from $r = 0$ to $r = 0.06$. Interestingly, as the localization radius increases the corresponding optimal inflation factor does too. At larger localization radii the filter makes more use of each observation leading to greater reduction in the posterior spread, which needs to be counterbalanced by increased inflation. The GP surrogates for analysis RMSE and for forecast metrics are qualitatively similar, as is the behavior of the hybrid filters. The optimal localization radii for the three filters are $L = 209$ (ESRF), $L = 279$ (SIR-ESRF), and $L = 238$ (BSIR-ESRF). These optimal values should not be over-interpreted, because the filters are not overly sensitive to the localization radius within the broad well that contains the optimal values. Nevertheless, the fact that the hybrids are able to use a larger localization radius might suggest that the particle filter resampling step is eliminating outliers that would otherwise lead to spurious long-range correlations.

All of the methods lead to under-dispersed ensembles in the sense that the ensemble spread is less than the RMSE. The forecast RMSE is 23% larger than the forecast spread for both ESRF and BSIR-ESRF, while it is 30% larger for SIR-ESRF. Forecast spread is here measured before inflation, but in every case the inflation is not enough to match the inflated spread to the forecast RMSE. The under-dispersion is worse for the analysis ensembles, with RMSE bigger than spread by 50% for ESRF and BSIR-ESRF, and by 80% for SIR-ESRF. This mismatch between spread and RMSE can be reduced by tuning the parameters (particularly by increasing the inflation), but only at the cost of increasing both the RMSE and the CRPS.

5.2 $N = 1200$

When the ensemble size is increased from 400 to 1200 the performance of the pure ESRF remains essentially flat, with only a minor improvement in analysis RMSE. This

Fig 5. The Gaussian Process (GP) model of analysis CRPS for the pure ESRF model at $N = 400$, showing analysis CRPS as a function of localization radius L and inflation ratio r. The left panel shows the mean of the GP and the right shows the standard deviation. The small gray squares indicate parameter values where experiments were run.

shows that for $N \geq 400$ the performance of the pure ESRF is limited primarily by the Gaussian approximation rather than by sampling errors. The optimal inflation parameter for ESRF reduces from $r = 0.026$ at $N = 400$ to $r = 0.015$ at $N = 1200$, and the optimal localization radius increases from $L = 209$ to $L = 250$. This is consistent with the intuition that as ensemble size increases less inflation and localization are needed to counteract sampling errors. The ensemble remains about as under-dispersed as at $N = 400$, with forecast RMSE 22% larger than forecast spread and analysis RMSE 42% larger than analysis spread.

The performance of the hybrid filters improves with increased ensemble size, with small improvements in CRPS and RMSE. The SIR-ESRF hybrid is now nearly indistinguishable from the pure ESRF, and the BSIR-ESRF hybrid outperforms both by 5-10% in CRPS and RMSE. It is not clear whether further improvements could be obtained by increasing the ensemble size, or whether the hybrids are already close to the true Bayesian posterior. No investigations have been performed at larger ensemble sizes due to the computational expense of optimizing parameters with very large ensembles.

Blurring of the observations enables the BSIR-ESRF hybrid to slightly out-perform the ESRF and the SIR-ESRF hybrid. The split parameter α for a fixed ESS threshold tends to be larger in the hybrid with blurred observations: at $N = 400$ the median α for SIR-ESRF is $10^{-3.14}$ while the median α for BSIR-ESRF is $10^{-2.91}$; at $N = 1200$ the median α for SIR-ESRF is $10^{-2.96}$ while the median α for BSIR-ESRF is $10^{-2.48}$. This suggests that for a fixed split α the blurring increases the ESS, but when the split α is instead chosen to produce a desired ESS the smoothing instead impacts which ensemble members are selected for resampling. Heuristically this can be explained as follows: The hybrid essentially decides a priori how many distinct ensemble members will remain after resampling (by fixing the ESS), so the only impact of the blurring will be on which ensemble members are eliminated and which are replicated. The effect of blurring is to trade improved assimilation performance at large scales for degraded performance at small scales; this trade is effective because predictability on longer time horizons comes from the large scales. Indeed, the RMSE of the analysis mean projected onto the large scale modes is better for BSIR-ESRF than the other two methods.

When the ensemble size increases from $N = 400$ to $N = 1200$ the optimal inflation for the SIR-ESRF hybrid decreases from $r = 0.06$ to $r = 0.04$, and the optimal localization radius increases from $L = 279$ to $L = 316$. The optimal ESS threshold is 757, although results are not overly sensitive for ESS thresholds in the range of 500 to 800. The optimal parameters of the BSIR-ESRF method are similar: $r = 0.02$, $L = 319$, and ESS threshold 642. Code and summary data associated with this section can be found in [58].

6 Conclusions

This paper has developed a hybrid particle ensemble Kalman filter targeting applications with medium nonlinearity, i.e. applications where the prior (forecast) distribution is very non-Gaussian but the posterior (analysis) distribution is close to Gaussian. It was argued in [38] that variational methods are more appropriate than the EnKF in this situation, because they approximate the posterior as Gaussian whereas

EnKF methods approximate the prior as Gaussian. The hybrid developed here is a pure ensemble approach for problems with medium nonlinearity, not requiring any variational optimization. The particle filter acts first and results in an intermediate distribution that is closer to Gaussian than the prior; this intermediate distribution is then presented to the EnKF, and matches more closely the Gaussian approximation inherent to the EnKF. The hybrid developed here is similar in spirit to previously-developed hybrids (e.g. [23, 24]). The main differences, besides the emphasis on medium nonlinearity, are the use of a serial square root filter and of a random resampling of the posterior ensemble to break the particle degeneracy introduced by the resampling step of the particle filter.

The hybrid SIR-ESRF developed here includes a resampling step that reduces the number of distinct ensemble members seen by the EnKF part of the hybrid (which is, in this case, a serial square root ESRF). The EnKF's performance is limited by sampling errors even in purely Gaussian problems, so reducing the number of distinct ensemble members used within the EnKF increases the sampling errors and can hurt performance. Our hybrid is configured such that the ESS in the particle filter step is specified a priori, and we find that the optimal ESS threshold for the hybrid needs to be at least as large as the ensemble size needed to obtain optimal performance in the pure EnKF. (ESRF performance stopped improving for ensemble sizes greater than 400, and the optimal ESS in the hybrid was between 500 and 800.) This leads one to expect that a larger ensemble size is required for the hybrid to outperform a pure EnKF, so that the particle filter component of the hybrid can effectively resample from the full ensemble size down to a size that is still large enough to obtain good EnKF performance. In problems where non-Gaussianity presents in the form of a few outliers in an otherwise nearly-Gaussian distribution, the hybrid will presumably need only a slightly larger ensemble size, so that it can eliminate outliers during the resampling step. But in problems where the forecast exhibits pathological non-Gaussianities such as multi-modality or strong curvature such as that seen in Fig 4, a much larger ensemble may be needed in order for the hybrid to outperform the pure EnKF. The SIR-ESRF hybrid did not achieve significant improvements over the pure ESRF in our experiments on the multiscale Lorenz-'96 model, but the BSIR-ESRF using smoothing of the innovations achieved limited improvements (5-10% improvement in CRPS and RMSE). This limited improvement compared to a pure ESRF may be a reflection of the fact that non-Gaussianity of the forecast is confined to a fairly low dimensional subspace associated with the leading singular vectors, i.e. the fastest directions of expansion along the system's attractor.

Supporting information

S1 Appendix Let $c_i(p)$ denote the observed mean CRPS for trial i with parameters p. It is reasonable to expect that the mean CRPS is a continuous latent function f of the filter parameters for fixed values of observed data, initial ensembles, random resamplings, and random rotations. But since these fixed values all vary in practice, we can view each f as a realization of a random field F. In this view, the quantities $c_i(p)$ are noisy observations of the random field's true mean $\overline{F}$. Our Bayesian optimizer seeks the minimizer of $\overline{F}$ using these noisy observations.

Let $\bar{c}_p$ be the mean CRPS observed over all assimilation trials that were run with parameters p. Then let $\sigma_{\bar{c}_p}$ be the empirical standard error of that mean, computed as the sample standard deviation divided by the square root of the number of trials. For convenience in setting hyperparameters of the Gaussian process model, we scale the search space to the unit cube and standardize the observations. The raw search spaces are hyperrectangles, so they are scaled in each coordinate in the obvious manner to arrive at a unit cube. To standardize the observations, we subtract the mean of the set

$\{\bar{c}_p\}$, for all parameter sets p previously evaluated in the experiment, and divide the result by the sample standard deviation $\sigma_{\bar{c}_p}$ of the same set. The raw standard errors $\sigma_{\bar{c}_p}$ are simultaneously divided by $\sigma_{\bar{c}_p}$ to preserve their validity in this standardized output space. We do not introduce new notation for these transformed quantities; the remainder of this section will treat c in the standardized output space and will treat values of p in the scaled parameter space.

In these scaled spaces, we form a surrogate model supposing that f_p depends on p as a Gaussian process

$$\mathcal{GP} \sim \mathcal{N}(0, k(p, p')). \tag{16}$$

We take $k(p, p')$ to be the Matérn covariance kernel

$$k(p_i, p_j) = \frac{\Theta_s 2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} d(p_i, p_j)\right) K_\nu \left(\sqrt{2\nu} \cdot d(p_i, p_j)\right), \tag{17}$$

where K_ν is the modified Bessel function of the second kind, and

$$d(\mathbf{p}_i, \mathbf{p}_j) = (\mathbf{p}_i - \mathbf{p}_j)^\top \boldsymbol{\Theta}_d^{-1} (\mathbf{p}_i - \mathbf{p}_j) \tag{18}$$

Here $\boldsymbol{\Theta}_d$ is a diagonal matrix of length scale hyperparameters. Each of the scalars on the diagonal of $\boldsymbol{\Theta}_d$ corresponds to a length scale of a feature in the space of scaled filter parameters, and each is endowed with a Gamma distribution prior $\Gamma(\lambda_L, r_L)$ with shape $\lambda_L = 6$ and rate $r_L = 3$. The factor Θ_s is another hyperparameter that controls the covariance function's overall scale, on which we also impose a Gamma distribution prior $\Gamma(\lambda_S, r_S)$ with shape $\lambda_S = 2$ and rate $r_S = 0.15$. We let the smoothness parameter $\nu = 5/2$ so that realizations are almost surely twice-differentiable. Marginalizing over the latent function f yields the posterior distribution with log density

$$\ln P\left(\mathbf{f} | \{\mathbf{p}_i\}, \Theta\right) = -\frac{1}{2}\mathbf{s}^\top (\mathbf{K} + \boldsymbol{\Xi})^{-1} \mathbf{s} - \frac{1}{2}\ln|\mathbf{K} + \boldsymbol{\Xi}| - \frac{N_p}{2}\ln(2\pi) \tag{19}$$

$$+ \sum_{j=1}^{N_p} \left[(\lambda_L - 1)\ln\left(\Theta_{L,j}\right) - r_L \Theta_{L,j} + \lambda_L \ln r_L - \ln\Gamma(\lambda_L)\right] \tag{20}$$

$$+ (\lambda_S - 1)\ln\left(\Theta_S\right) - r_S \Theta_{S,j} + \lambda_S \ln r_S - \ln\Gamma(\lambda_S), \tag{21}$$

where $K_{ij} = k(p_i, p_j)$ is a covariance matrix. The equation above obtains by adding the log-likelihood of our hyperparameter priors to Equation 2.30 of [66]. The GP surrogate is then fit to the rescaled data by maximizing the log-density Eq 19 using many restarts of the L-BFGS-B method [67] to arrive at the maximum a posteriori (MAP) estimator. Finally, a batch of candidate arms is generated that approximately optimizes the batched noisy expected improvement acquisition function [68] on the MAP estimator. Batch sizes varied between 1 and 32 depending on computational resources available at the time.

References

1. Evensen G. Data Assimilation: The Ensemble Kalman Filter. Springer; 2009.
2. Mandel J, Cobb L, Beezley JD. On the convergence of the ensemble Kalman filter. Appl Math. 2011;56(6):533–541.
3. Gordon NJ, Salmond DJ, Smith AF. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: IEEE Proceedings F (Radar and Signal Processing). vol. 140. IET; 1993. p. 107–113.

4. Crisan D, Doucet A. A survey of convergence results on particle filtering methods for practitioners. IEEE T Signal Proces. 2002;50(3):736–746.

5. Law K, Stuart A, Zygalakis K. Data assimilation. Springer: Cham, Switzerland; 2015.

6. Bengtsson T, Bickel P, Li B. In: Nolan D, Speed T, editors. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. vol. Volume 2 of Collections. Beachwood, Ohio, USA: Institute of Mathematical Statistics; 2008. p. 316–334. Available from: http://dx.doi.org/10.1214/193940307000000518.

7. Snyder C, Bengtsson T, Bickel P, Anderson J. Obstacles to high-dimensional particle filtering. Mon Wea Rev. 2008;136(12):4629–4640.

8. Snyder C, Bengtsson T, Morzfeld M. Performance bounds for particle filters using the optimal proposal. Mon Wea Rev. 2015;143(11):4750–4761.

9. Chorin AJ, Tu X. Implicit sampling for particle filters. Proc Natl Acad Sci (USA). 2009;106(41):17249–17254.

10. Chorin AJ, Tu X. An iterative implementation of the implicit nonlinear filter. ESAIM-Math Model Num. 2012;46(3):535–543.

11. Chorin A, Morzfeld M, Tu X. Implicit particle filters for data assimilation. Comm App Math Com Sc. 2010;5(2):221–240.

12. Van Leeuwen PJ. Particle filtering in geophysical systems. Mon Wea Rev. 2009;137(12):4089–4114.

13. Ades M, Van Leeuwen PJ. An exploration of the equivalent weights particle filter. Quart J Roy Meteor Soc. 2013;139(672):820–840.

14. Ades M, Van Leeuwen PJ. The equivalent-weights particle filter in a high-dimensional system. Quart J Roy Meteor Soc. 2015;141(687):484–503.

15. Zhu M, van Leeuwen PJ, Amezcua J. Implicit equal-weights particle filter. Quart J Roy Meteor Soc. 2016;142(698):1904–1919. doi:https://doi.org/10.1002/qj.2784.

16. Skauvold J, Eidsvik J, van Leeuwen PJ, Amezcua J. A revised implicit equal-weights particle filter. Quart J Roy Meteor Soc. 2019;145(721):1490–1502. doi:https://doi.org/10.1002/qj.3506.

17. Robinson G, Grooms I, Kleiber W. Improving particle filter performance by smoothing observations. Mon Wea Rev. 2018;146(8):2433–2446.

18. Rebeschini P, Van Handel R. Can local particle filters beat the curse of dimensionality? Ann Appl Probab. 2015;25(5):2809–2866.

19. Penny SG, Miyoshi T. A local particle filter for high-dimensional geophysical systems. Nonlinear Proc Geoph. 2016;23(6):391–405.

20. Poterjoy J. A localized particle filter for high-dimensional nonlinear systems. Mon Wea Rev. 2016;144(1):59–76.

21. Pulido M, van Leeuwen PJ. Sequential Monte Carlo with kernel embedded mappings: The mapping particle filter. J Comput Phys. 2019;.

22. Pinheiro FR, van Leeuwen PJ, Geppert G. Efficient nonlinear data assimilation using synchronization in a particle filter. Quart J Roy Meteor Soc. 2019;145(723):2510–2523.

23. Frei M, Künsch HR. Bridging the ensemble Kalman and particle filters. Biometrika. 2013;100(4):781–800.

24. Chustagulprom N, Reich S, Reinhardt M. A hybrid ensemble transform particle filter for nonlinear and spatially extended dynamical systems. SIAM/ASA J Uncertainty Quantification. 2016;4(1):592–608.

25. Papadakis N, Mémin É, Cuzol A, Gengembre N. Data assimilation with the weighted ensemble Kalman filter. Tellus A. 2010;62(5):673–697.

26. Slivinski L, Spiller E, Apte A, Sandstede B. A hybrid particle–ensemble Kalman filter for Lagrangian data assimilation. Mon Wea Rev. 2015;143(1):195–211.

27. Morzfeld M, Hodyss D, Poterjoy J. Variational particle smoothers and their localization. Quart J Roy Meteor Soc. 2018;144(712):806–825.

28. Bertino L, Evensen G, Wackernagel H. Sequential data assimilation techniques in oceanography. International Statistical Review. 2003;71(2):223–241.

29. Bocquet M, Pires CA, Wu L. Beyond Gaussian statistical modeling in geophysical data assimilation. Mon Wea Rev. 2010;138(8):2997–3023.

30. Zhou H, Gomez-Hernandez JJ, Franssen HJH, Li L. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. Adv Water Resour. 2011;34(7):844–864.

31. Brankart JM, Testut CE, Béal D, Doron M, Fontana C, Meinvielle M, et al. Towards an improved description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations. Ocean Science. 2012;8(2):121.

32. Simon E, Bertino L. Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model. J Marine Syst. 2012;89(1):1–18.

33. Anderson JL. A non-Gaussian ensemble filter update for data assimilation. Mon Wea Rev. 2010;138(11):4186–4198.

34. Metref S, Cosme E, Snyder C, Brasseur P. A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation. Nonlinear Proc Geoph. 2014;21:869–885.

35. Anderson JL. A nonlinear rank regression method for ensemble Kalman filter data assimilation. Mon Wea Rev. 2019;147(8):2847–2860.

36. Anderson JL. A Marginal Adjustment Rank Histogram Filter for Non-Gaussian Ensemble Data Assimilation. Mon Wea Rev. 2020;148(8):3361–3378.

37. Bocquet M. Ensemble Kalman filtering without the intrinsic need for inflation. Nonlinear Proc Geoph. 2011;18(5):735–750. doi:10.5194/npg-18-735-2011.

38. Morzfeld M, Hodyss D. Gaussian approximations in filters and smoothers for data assimilation. Tellus A. 2019;71(1):1–27.

39. Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. J Amer Stat Assoc. 1986;81(393):82–86.

40. Evensen G. Sampling strategies and square root analysis schemes for the EnKF. Ocean dynamics. 2004;54(6):539–560.

41. Leeuwenburgh O, Evensen G, Bertino L. The impact of ensemble filter definition on the assimilation of temperature profiles in the tropical Pacific. Quart J Roy Meteor Soc. 2005;131(613):3291–3300.

42. Sakov P, Oke PR. Implications of the form of the ensemble transformation in the ensemble square root filters. Mon Wea Rev. 2008;136(3):1042–1053.

43. Lei J, Bickel P. A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. Mon Wea Rev. 2011;139(12):3964–3973.

44. Tödter J, Ahrens B. A second-order exact ensemble square root filter for nonlinear data assimilation. Mon Wea Rev. 2015;143(4):1347–1367.

45. Grooms I, Lee Y. A framework for variational data assimilation with superparameterization. Nonlinear Proc Geoph. 2015;22(5):601–611.

46. Doucet A, De Freitas N, Gordon N. An Introduction to Sequential Monte Carlo methods. In: Sequential Monte Carlo methods in practice. Springer; 2001. p. 3–14.

47. Kitagawa G. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. J Comput Graph Stat. 1996;5(1):1–25.

48. Reich S. A nonparametric ensemble transform method for Bayesian inference. SIAM J Sci Comput. 2013;35(4):A2013–A2024.

49. Acevedo W, de Wiljes J, Reich S. Second-order accurate ensemble transform particle filters. SIAM J Sci Comput. 2017;39(5):A1834–A1850.

50. Whitaker JS, Hamill TM. Ensemble data assimilation without perturbed observations. Mon Wea Rev. 2002;130(7):1913–1924.

51. Anderson JL, Anderson SL. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon Wea Rev. 1999;127(12):2741–2758.

52. Burgers G, Jan van Leeuwen P, Evensen G. Analysis scheme in the ensemble Kalman filter. Mon Wea Rev. 1998;126(6):1719–1724.

53. Houtekamer PL, Mitchell HL. Data assimilation using an ensemble Kalman filter technique. Mon Wea Rev. 1998;126(3):796–811.

54. Robinson G, Grooms I. A Fast Tunable Blurring Algorithm for Scattered Data. SIAM Journal on Scientific Computing. 2020;42(4):A2281–A2299.

55. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc. 2007;102:359–378.

56. Hersbach H. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather Forecast. 2000;15:559–570.

57. Bengtsson T, Snyder C, Nychka D. Toward a nonlinear ensemble filter for high-dimensional systems. J Geophys Res: Atmospheres. 2003;108(D24).

58. Grooms I. iangrooms/RG_PONE_Hybrids: Zenodo, take 2; 2020. Available from: `https://doi.org/10.5281/zenodo.4394312`.

59. Lorenz E. Predictability: A problem partly solved. In: Proceedings of Seminar on Predicability. vol. 1. ECMWF. Reading, UK; 1996. p. 1–18.

60. Lorenz E. Predictability: A problem partly solved. In: Palmer T, Hagedorn R, editors. Predictability of Weather and Climate. Cambridge University Press; 2006. p. 40–58.

61. Grooms I. Analog ensemble data assimilation and a method for constructing analogs with variational autoencoders. Quart J Roy Meteor Soc;147(734):139–149. doi:10.1002/qj.3910.

62. Lorenz EN, Emanuel KA. Optimal sites for supplementary weather observations: Simulation with a small model. J Atmos Sci. 1998;55(3):399–414.

63. Owen AB. Scrambling Sobol and Niederreiter–Xing points. J Complexity. 1998;14(4):466–489.

64. Anderson JL. Reducing correlation sampling error in ensemble Kalman filter data assimilation. Mon Wea Rev. 2016;144(3):913–925.

65. El Gharamti M, Raeder K, Anderson J, Wang X. Comparing Adaptive Prior and Posterior Inflation for Ensemble Filters Using an Atmospheric General Circulation Model. Mon Wea Rev. 2019;147(7):2535–2553.

66. Williams CK, Rasmussen CE. Gaussian processes for machine learning. vol. 2. MIT press Cambridge, MA; 2006.

67. Byrd RH, Lu P, Nocedal J, Zhu C. A limited memory algorithm for bound constrained optimization. SIAM J Sci Comput. 1995;16(5):1190–1208.

68. Letham B, Karrer B, Ottoni G, Bakshy E, et al. Constrained Bayesian optimization with noisy experiments. Bayesian Analysis. 2019;14(2):495–519.